Title: Heterologous production of caffeic acid from tyrosine in *Escherichia coli*

Author: J.L. Rodrigues R.G. Araújo K.L.J. Prather L.D. Kluskens L.R. Rodrigues

PII: S0141-0229(15)00002-2
DOI: http://dx.doi.org/doi:10.1016/j.enzmictec.2015.01.001
Reference: EMT 8718

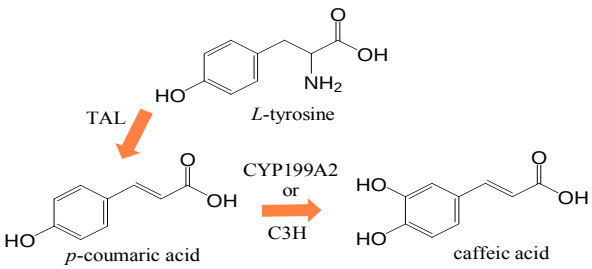
To appear in: *Enzyme and Microbial Technology*

Received date: 16-10-2014
Revised date: 30-12-2014
Accepted date: 6-1-2015

Please cite this article as: Rodrigues JL, Araújo RG, Prather KLJ, Kluskens LD, Rodrigues LR, Heterologous production of caffeic acid from tyrosine in *Escherichia coli*, *Enzyme and Microbial Technology* (2015), http://dx.doi.org/10.1016/j.enzmictec.2015.01.001

1

2



Highlights

- *Escherichia coli* was engineered for the production of caffeic acid.
- Tyrosine ammonia lyase (TAL) converted 3 mM of tyrosine to 2.62 mM *p*-coumaric acid.
- TAL and 4-coumarate 3-hydroxylase (C3H) converted tyrosine in 1 mM caffeic acid.
- This is the first study that shows caffeic acid production using TAL and CYP199A2.
- TAL and CYP199A2 converted tyrosine in 1.56 mM caffeic acid.

Heterologous production of caffeic acid from tyrosine in *Escherichia coli*

J. L. Rodrigues[a,c], R. G. Araújo[a], K. L. J. Prather[b,c], L.D. Kluskens[a], L.R. Rodrigues[a,c*]

[a]*Centre of Biological Engineering, University of Minho, 4710 - 057 Braga, Portugal*

[b]*Department of Chemical Engineering, Synthetic Biology Engineering Research Center (SynBERC) Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

[c]*MIT-Portugal Program, Cambridge, MA and Lisbon, Portugal*

J. L. Rodrigues – joana.joanalucia@deb.uminho.pt
R. G. Araújo - rafa.gomes.ar@gmail.com
K. L. J. Prather - kljp@mit.edu
L.D. Kluskens – kluskens@deb.uminho.pt
L.R. Rodrigues - lrmr@deb.uminho.pt

[*]Corresponding author: Lígia R. Rodrigues
Address: Centre of Biological Engineering, University of Minho, 4710 - 057 Braga, Portugal
Tel: (+351) 253 604 401
Fax: (+351) 253 604 429
E-mail address: lrmr@deb.uminho.pt

Abstract

Caffeic acid is a plant secondary metabolite and its biological synthesis has attracted increased attention due to its beneficial effects on human health. In this study, *Escherichia coli* was engineered for the production of caffeic acid using tyrosine as the initial precursor of the pathway. The pathway design included tyrosine ammonia lyase (TAL) from *Rhodotorula glutinis* to convert tyrosine to *p*-coumaric acid and 4-coumarate 3-hydroxylase (C3H) from *Saccharothrix espanaensis* or cytochrome P450 CYP199A2 from *Rhodopseudomonas palustris* to convert *p*-coumaric acid to caffeic acid. The genes were codon-optimized and different combinations of plasmids were used to improve the titer of caffeic acid. TAL was able to efficiently convert 3 mM of tyrosine to *p*-coumaric acid with the highest production obtained being 2.62 mM (472 mg/L). CYP199A2 exhibited higher catalytic activity towards *p*-coumaric acid than C3H. The highest caffeic acid production obtained using TAL and CYP199A2 and TAL and C3H was 1.56 mM (280 mg/L) and 1 mM (180 mg/L), respectively. This is the first study that shows caffeic acid production using CYP199A2 and tyrosine as the initial precursor. This study suggests the possibility of further producing more complex plant secondary metabolites like flavonoids and curcuminoids.

Keywords: tyrosine; *p*-coumaric acid; caffeic acid; *E. coli*; biosynthesis; synthetic biology

1. Introduction

Caffeic acid is a natural phenolic compound derived from the phenylpropanoid pathway in plants. Phenylpropanoic acids, especially caffeic acid, have attracted an increased attention owing to their valuable properties, including antioxidant [1], anti-inflammatory [2], anticancer [3], antiviral [4], antidiabetic [5] and antidepressive [6]. Due to its pharmaceutical applications there is an established market for caffeic acid production. Caffeic acid is mainly obtained by extraction from plants and these extraction methods include high temperatures, treatment with petroleum and solvent extraction (methanol, ethyl acetate), thus being energy-intensive and environmentally unfriendly [7]. Moreover, the yields obtained are low because it accumulates at low levels in plant cells. To meet the phenylpropanoids market needs, it is imperative to look for new green and environmental production methods to replace the traditional extraction one. One way to achieve high levels of caffeic acid is the production via engineered microorganisms, such as *Escherichia coli* [8-10].

Natural caffeic acid production starts with the deamination of the amino acid phenylalanine which is converted to cinnamic acid by phenylalanine ammonia lyase (PAL) (Fig. 1). Then, cinnamic acid is converted ~~in~~ to *p*-coumaric acid by cinnamate-4-hydroxylase (C4H), and caffeic acid is obtained from *p*-coumaric acid using 4-coumarate 3-hydroxylase (C3H). Tyrosine can also be used as a precursor as some PALs also have tyrosine ammonia lyase (TAL) activity [11, 12]. Since tyrosine already possesses a 4-hydroxyl group, its use is advantageous as it can be directly converted to *p*-coumaric acid, thus decreasing the number of steps to produce caffeic acid. Also, using TAL, the C4H enzyme that is essential

for caffeic acid production in plants and has not yet been successfully expressed in prokaryotic organisms [13] is not needed anymore.

In the last decade several efforts have been conducted to produce caffeic acid using microorganisms. Berner et al. [14] identified *sam8* and *sam5* genes, encoding TAL and C3H involved in caffeic acid biosynthesis in the actinomycete *Saccharothrix espanaensis* and expressed them in *Streptomyces fradiae* XKS. After that, these enzymes were used to produce caffeic acid in *E. coli* (Table 1). Other bacterial TAL (*Rhodobacter capsulatus* and *Rhodotorula glutinis*) were also used for the production of *p*-coumaric acid and hydroxyphenylacetate 3-hydroxylase (4HPA3H) from *E. coli* or *Pseudomonas aeruginosa*, and cytochrome P450 CYP199A2 from *Rhodopseudomonas palustris* was proved to convert *p*-coumaric acid to ~~in~~ caffeic acid with a high yield.

In this study, we describe the production of caffeic acid from tyrosine or *p*-coumaric acid. To convert tyrosine to ~~in~~ *p*-coumaric acid we used TAL from *R. glutinis*. *p*-coumaric acid was converted to ~~in~~ caffeic acid using C3H from *S. espanaensis* or CYP199A2 from *R. palustris*. Although all these genes have previously been used in the caffeic acid production, it is important to mention that, as far as we know, the caffeic acid production from tyrosine using CYP199A2 in the pathway has never been attempted. Additionally, in the current study, the titers of caffeic acid obtained using TAL and C3H genes were higher than the ones reported in other studies using the same genes. These high yields were obtained by using different combinations of plasmids and genetic arrangements.

106 Furthermore, it is important to mention that the caffeic acid pathway can be further used to
107 produce other products of the phenylpropanoid pathway with high added-value like
108 flavonoids and curcuminoids.

Table 1. Caffeic acid production in *Escherichia coli.* Genes/organisms used in the caffeic acid biosynthetic pathway, fermentation conditions and yields.

Genes – Organisms[a]	Fermentation Conditions	Titer (mg/L)[b]	Reference
TAL - *S. espanaensis* C3H - *S. espanaensis*	First induction in LB at 37 ºC. After 5 h cells were transferred to M9 medium (glucose) at 26 ºC for 36 h	(not mentioned)	Choi et al. [15]
TAL - *R. capsulatus* 4HPA3H - *E. coli* Δ*tyrR*; *tyrA*^{fbr}-*ppsA-tktA-aroG*^{fbr}	M9 medium (glucose and glycerol) at 37 ºC. After induction shake flasks were transferred to 30 ºC for 48 h	50.2 (TYR strain) 12.1 (wild-type)	Lin and Yan [16]
TAL and opTAL - *S. espanaensis* C3H - *S. espanaensis* Δ*tyrR*; *tyrA*^{fbr}- *aroG*^{fbr}	First induction in LB at 37 ºC. After 5 h cells were harvested and transferred to M9 medium (glucose) at 26 ºC for 36 h in shake flasks	150 (TAL, TYR strain) 40 (opTAL, TYR strain) 42 (TAL, wild-type) 14 (opTAL, wild-type)	Kang et al. [17]
CYP199A2 (wild-type and mutant F185L) - *R. palustris* Pdr - *Pseudomonas putida* Pux - *R. palustris*	Potassium phosphate buffer (glucose or glycerol) at 30 ºC in shake flasks for 24 h. 20 mM *p*-coumaric acid was added.	510 (wild-type) 2800 (mutant)	Furuya et al. [8]
TAL - *R. glutinis* C3H - *S. espanaensis* Δ*pheA* Δ*tyrR*; *tyrA*^{fbr}-*aroG*^{fbr}	MOPS medium (glucose or xylose), synthetic medium (glucose or xylose) or LB (glucose or xylose) at 37 ºC for 72 h in test tubes; ~~S~~ Synthetic medium (glucose) at 37 ºC for 7 days in bioreactor	88 (LB, glucose, test tube) 106 (bioreactor)	Zhang and Stephanopoulos [18][c]
TAL - *R. glutinis* 4HPA3H - *E. coli* Δ*pheA*; *tyrA*^{fbr}-*ppsA-tktA-aroG*^{fbr}	M9 medium (yeast extract and glycerol, or glucose and glycerol) at 37 ºC for 72 h in shake flasks. 20 mM *p*-coumaric acid fed 3 h after induction and 3 mM added afterward (wild-type case)	3820 (wild-type) 766.68 (TYR strain)	Huang et al. [10]
4HPA3H - *P. aeruginosa*	Potassium phosphate buffer (glucose or glycerol) at 30 ºC for 24 h in shake flasks. 20 mM of *p*-coumaric acid was added 4 ~~3~~ times	10200	Furuya and Kino [9]

[a]In some studies a tyrosine overproducing strain was used in which: *tyrR* (tyrosine repressor) and *pheA* (chorismate mutase / prephenate dehydratase) were deleted (Δ*tyrR* and Δ*pheA*) to direct the pathway only to tyrosine production; *ppsA* (PEP synthase) and *tktA* (transkelotase) were overexpressed to increase the availability of the two main precursors of aromatic amino acids biosynthesis; and *tyrA* (chorismate mutase / prephenate dehydrogenase) and *aroG* (3-deoxy-D-arabino-heptulosonate 7-phosphate synthase) gene sequences were modified and the feedback inhibition-resistant derivatives (*tyrA*^{fbr} and *aroG*^{fbr}) were overexpressed to remove the transcriptional control mediated by tyrosine and phenylalanine. OpTAL means codon-optimized TAL. [b]TYR strain – tyrosine overproducing strain; [c]An alternative route through coumaroyl-CoA and caffeoyl-CoA (Fig. 1) using 4CL from *P. crispus* and *E. coli* endogenous thiosterase was tested to increase the yield but it was not successful.

2. Materials and Methods

2.1 Bacterial strains, plasmids and chemicals

E. coli ElectroMAX™ DH10B competent cells (Invitrogen/Life Technologies, Carlsbad, CA, USA) were used for molecular cloning and vector propagation. *E. coli* K-12 MG1655(DE3) [19] was used as the host for the expression of genes under the control of the T7 promoter.

P. putida JCM 6157 strain (ATCC 17453, Manassas, VA, USA) and *R. palustris* CGA009 gDNA (ATCC BAA-98) were used to amplify the *pdr* and *pux* genes. The characteristics of all the strains and plasmids used in this study are described in Table 2.

Restriction, ligation and Q5 enzymes (NEB, Ipswich, MA, USA), QIAprep Spin Miniprep Kit (Qiagen, Germantown, MD, USA), DNA Clean and Concentrator and Gel DNA Recovery Kits (Zymo Research, Orange, CA, USA) and Wizard® Genomic DNA Purification Kit (Promega, Madison, WI, USA) were used according to the instructions provided by the manufacturers.

L-Tyrosine, *p*-coumaric acid and caffeic acid were purchased from Sigma-Aldrich (Steinheim, Germany), isopropyl β-D-thiogalactopyranoside (IPTG) and Luria-Bertani (LB) medium from NZYTech (Lisbon, Portugal) and anhydrotetracycline (aTc) from Acros (Geel, Belgium). Glucose (Acros), Na_2HPO_4 (Scharlau, Sentmenat, Spain), $MgSO_4$, KH_2PO_4 (Riel-deHaën, Seelze, Germany), NH_4Cl, NaCl, $CaCO_3$ (Panreac, Barcelona, Spain) and thiamine (Fisher Scientific, Loughborough, UK) were used to prepare the M9 modified salt medium. The following mineral traces and vitamins were supplemented to the M9 medium: $FeCl_3$, $ZnCl_2$, $CoCl_2$, $CuCl_2$, nicotinic acid (Riedel-deHaën), $NaMoO_4$, H_2BO_3, pyridoxine, biotin, folic acid (Merck), riboflavin and pantothenic acid (Sigma

139 Aldrich). Ampicillin (Applichem, Darmstadt, Germany), chloramphenicol, kanamycin

140 (NZYtech) and spectinomycin (Panreac) were used when necessary.

141

142 **Table 2**. Bacterial strains and plasmids used in this study

Strains	**Relevant Genotype**	**Source**
E. coli Electromax DH10B	*F⁻ mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*Δ*M15* Δ*lacX*74 *recA*1 *endA*1 *araD*139 Δ(*ara*, *leu*)7697 *galU galK λ⁻ rpsL nupG*	Invitrogen/Life Technologies
E. coli K-12 MG1655(DE3)	*F⁻ λ⁻ ilvG⁻ rfb⁻*50 *rph⁻*1 λ(DE3)	[19]
P. putida JCM 6157		ATCC 17453
Genomic DNA		**Source**
R. palustris CGA009		ATCC BAA-98D-5
Plasmids	**Construct**	**Source**
pETDuet-1	ColE1(pBR322) *ori*, *lacI*, double T7*lac*, Amp^R	Novagen
pCDFDuet-1	CloDF13 *ori*, *lacI*, double T7*lac*, $Strep^R$	Novagen
pRSFDuet-1	RSF *ori*, *lacI*, double T7*lac*, Kan^R	Novagen
pKVS45	*p*15*A ori*, tetR, P_{tet}, Amp^R	[20]
pUC57_TAL	pUC57 carrying codon-optimized TAL from *R. glutinis*	GenScript
pUC57_C3H	pUC57 carrying codon-optimized C3H from *S. espanaensis*	GenScript
pUC57_CYP199A2	pUC57 carrying codon-optimized CYP199A2 from *R. palustris*	GenScript
pETDuet_TAL	pETDuet-1 carrying codon-optimized TAL from *R. glutinis*	This study
pETDuet_C3H	pETDuet-1 carrying codon-optimized C3H from *S. espanaensis*	This study
pETDuet_TAL_C3H	pETDuet_TAL carrying codon-optimized C3H from *S. espanaensis*	This study
pCDFDuet_TAL	pCDFDuet-1 carrying codon-optimized TAL from *R. glutinis*	This study
pCDFDuet_C3H	pCDFDuet-1 carrying codon-optimized C3H from *S. espanaensis*	This study
pCDFDuet_CYP	pCDFDuet-1 carrying codon-optimized CYP199A2 from *R. palustris*	This study
pCDFDuet_CYP (+7aa)	pCDFDuet-1 carrying codon-optimized CYP199A2 from *R. palustris* with the first 7 amino acids	This study
pCDFDuet_TAL_CYP	pCDFDuet_TAL carrying codon-optimized CYP199A2 from *R. palustris*	This study
pCDFDuet_TAL_CYP (+7aa)	pCDFDuet_TAL carrying CYP199A2 from *R. palustris* with the first 7 amino acids	This study
pCDFDuet_TAL_CYP_op	pCDFDuet-1 carrying codon-optimized TAL from *R. glutinis* and CYP199A2 from *R. palustris* in an operon	This study
pCDFDuet_TAL_CYP(+7aa)_op	pCDFDuet-1 carrying codon-optimized TAL from *R. glutinis* and CYP199A2 from *R. palustris* with the first 7 amino acids in an operon	This study
pKVS45_TAL	pKVS45 carrying codon-optimized TAL from *R.*	This study

	glutinis	
pKVS45_C3H	pKVS45 carrying codon-optimized C3H from *S. espanaensis*	This study
pRSFDuet_TAL	pRSFDuet-1 carrying codon-optimized TAL from *R. glutinis*	This study
pRSFDuet_C3H	pRSFDue-1t carrying codon-optimized C3H from *S. espanaensis*	This study
pKVS45_Pdr_Pux_op	pKVS45 carrying Pdr from *P. putida* and Pux from *R. palustris* in an operon	This study
pETDuet_Pdr_Pux_op	pETDuet-1 carrying Pdr from *P. putida* and Pux from *R. palustris* in an operon	This study

2.2 Codon-optimization and synthesis of TAL, C3H and CYP199A2

TAL, C3H and CYP199A2 genes were codon-optimized for *E. coli*, synthesized and cloned in pUC57 vector by GenScript (Piscataway, NJ, USA). In addition to codon-optimization, the phenylalanine residue at 185 position (F185) of CYP199A2 was replaced by leucine (F185L)[8]. The DNA sequences of the codon-optimized genes are provided in Table 3. ~~Supplementary Material (Table S1)~~.

158 **Table 3.** Gene sequences of TAL, C3H and CYP199A2 with codon optimization.

Gene	Sequence
TAL (*Rhodotorula glutinis*)	ATGGCTCCGCGTCCGACCTCGCAATCCCAAGCTCGCACCTGCCCGACCACCCAAGTTACCCAAGTTGACATCGTTGAAAAAA TGCTGGCGGCGCCGACCGATTCGACGCTGGAACTGGACGGCTATAGCCTGAACCTGGGTGATGTGGTTTCTGCAGCACGTAA AGGTCGTCCGGTGCGTGTTAAAGATTCAGACGAAATTCGCTCGAAAATCGATAAAAGCGTGGAATTTCTGCGTAGCCAGCTG AGCATGTCTGTTTACGGCGTCACCACGGGTTTCGGCGGTTCAGCCGATACCCGCACGGAAGACGCCATTTCGCTGCAGAAAG CACTGCTGGAACATCAACTGTGCGGCGTGCTGCCGAGCTCTTTTGATAGCTTCCGCCTGGGCCGTGGTCTGGAAAACTCTCTG CCGCTGGAAGTCGTGCGTGGTGCAATGACCATCCGTGTTAATTCCCTGACGCGCGGTCATTCAGCTGTCCGTCTGGTTGTCCT GGAAGCGCTGACCAACTTTCTGAATCACGGTATTACGCCGATCGTGCCGCTGCGTGGTACCATTAGTGCATCCGGTGATCTG AGCCCGCTGTCTTATATTGCAGCTGCGATCTCTGGCCACCCGGACAGTAAAGTTCATGTGGTTCACGAGGGTAAAGAAAAAA TCCTGTACGCCCGTGAAGCTATGGCGCTGTTCAACCTGGAACCGGTCGTGCTGGGCCCGAAAGAAGGCCTGGGTCTGGTGAA TGGTACGGCTGTTTCAGCGTCGATGGCCACCCTGGCACTGCATGATGCCCACATGCTGAGCCTGCTGAGCCAGTCTCTGACC GCGATGACGGTCGAAGCGATGGTGGGCCATGCAGGTAGCTTTCATCCGTTCCTGCACGATGTGACCCGTCCGCACCCGACGC AGATTGAAGTTGCAGGCAACATCCGCAAACTGCTGGAAGGTAGCCGTTTTGCGGTGCATCACGAAGAAGAAGTGAAAGTGA AAGATGACGAAGGCATTCTGCGCCAGGATCGTTATCCGCTGCGTACCAGTCCGCAATGGCTGGGTCCGCTGGTCTCCGACCT GATTCATGCCCACGCAGTGCTGACCATCGAAGCGGGTCAGAGTACCACGGATAACCCGCTGATTGACGTGGAAAATAAAAC CTCTCATCACGGCGGTAACTTTCAAGCCGCAGCTGTTGCCAATACGATGGAAAAAACGCGCCTGGGCCTGGCACAGATCGGT AAACTGAATTTCACCCAACTGACGGAAATGCTGAACGCAGGCATGAATCGTGGTCTGCCGAGCTGCCTGGCAGCAGAAGAT CCGAGTCTGTCCTATCATTGTAAAGGCCTGGACATTGCAGCTGCGGCCTACACCTCTGAACTGGGTCATCTGGCGAACCCGG TTACCACGCACGTCCAGCCGGCTGAAATGGCGAACCAAGCCGTGAATTCCCTGGCACTGATCTCAGCTCGTCGCACCACGGA ATCGAATGATGTCCTGAGCCTGCTGCTGGCGACCCATCTGTATTGTGTTCTGCAGGCTATTGACCTGCGCGCGATCGAATTTG AATTCAAAAAACAGTTTGGCCCGGCTATTGTGAGCCTGATCGATCAACACTTCGGCTCTGCCATGACCGGTAGTAACCTGCG TGACGAACTGGTGGAAAAAGTTAATAAAACGCTGGCCAAACGCCTGGAACAGACCAACAGTTACGATCTGGTGCCGCGTTG GCATGACGCATTTTCCTTCGCAGCTGGTACGGTTGTCGAAGTTCTGAGTTCCACCTCACTGTCGCTGGCGGCCGTCAATGCCT GGAAAGTGGCAGCTGCGGAAAGTGCAATTTCCCTGACCCGCCAAGTGCGTGAAACGTTTTGGTCAGCAGCATCGACGTCATC GCCGGCACTGAGCTATCTGTCTCCGCGCACCCAAATTCTGTACGCTTTTGTTCGTGAAGAACTGGGCGTCAAAGCGCGTCGC GGCGATGTTTTCCTGGGTAAACAGGAAGTGACCATCGGTAGTAATGTTTCCAAAATCTATGAAGCTATCAAAAGCGGTCGTA TCAATAATGTGCTGCTGAAAATGCTGGCATAA

159

160

161

Table 3. Gene sequences of TAL, C3H and CYP199A2 with codon optimization (*continuation*).

Gene	Sequence
C3H (*Saccharothrix espanaensis*)	ATGACGATTACCTCTCCGGCACCGGCTGGTCGCCTGAACAATGTCCGCCCGATGACGGGTGAAGAATACCTGGAATCCCTGC GTGACGGCCGTGAAGTGTATATTTACGGTGAACGCGTCGATGACGTGACCACGCATCTGGCGTTCCGCAACAGCGTTCGTTC TATCGCCCGCCTGTATGATGTCCTGCACGACCCGGCATCCGAAGGTGTTCTGCGTGTCCCGACCGATACGGGTAATGGTGGTT TTACCCATCCGTTTTTCAAAACGGCGCGTAGCTCTGAAGACCTGGTGGCGGCCCGTGAAGCCATTGTCGGCTGGCAACGCCT GGTGTATGGCTGGATGGGTCGTACCCCGGATTACAAAGCAGCGTTTTTCGGTACGCTGGACGCTAACGCGGAATTTTATGGC CCGTTCGAAGCCAATGCACGTCGCTGGTATCGTGATGCACAGGAACGCGTTCTGTACTTCAACCATGCTATCGTTCACCCGCC GGTCGATCGTGACCGTCCGGCTGATCGTACCGCAGACATTTGCGTCCATGTGGAAGAAGAAACGGATTCAGGCCTGATCGTG TCGGGTGCCAAAGTGGTTGCAACCGGTTCTGCTATGACGAACGCGAATCTGATTGCCCACTATGGTCTGCCGGTTCGCGATA AAAAATTTGGCCTGGTGTTCACCGTTCCGATGAACAGTCCGGGTCTGAAACTGATCTGTCGTACCTCCTATGAACTGATGGTG GCCACGCAGGGCTCACCGTTTGATTACCCGCTGAGTTCCCGCCTGGATGAAAATGACAGCATTATGATCTTTGATCGTGTTCT GGTCCCGTGGGAAAACGTTTTCATGTACGACGCAGGCGCGGCCAATAGCTTTGCTACCGGCTCTGGTTTCCTGGAACGCTTTA CCTTCCATGGTTGCACGCGTCTGGCAGTGAAACTGGATTTTATTGCAGGCTGTGTTATGAAAGCTGTGGAAGTTACCGGCACC ACGCACTTCCGCGGTGTTCAGGCGCAAGTCGGCGAAGTGCTGAACTGGCGTGATGTCTTTTGGGGTCTGTCGGACGCTATGG CGAAAAGTCCGAATTCCTGGGTGGGCGGTAGCGTTCAGCCGAACCTGAATTATGGCCTGGCCTACCGCACCTTTATGGGCGT GGGTTATCCGCGTATTAAAGAAATTATCCAGCAAACGCTGGGCTCTGGTCTGATCTACCTGAACTCATCGGCAGCTGATTGG AAAAATCCGGACGTTCGCCCGTATCTGGATCGTTACCTGCGCGGCAGTCGTGGTATTCAGGCAATCGATCGTGTCAAACTGC TGAAACTGCTGTGGGACGCAGTGGGTACCGAATTCGCAGGTCGTCATGAACTGTATGAACGCAACTACGGCGGTGATCACG AAGGTATTCGTGTGCAGACCCTGCAAGCCTATCAGGCAAATGGTCAAGCGGCCGCACTGAAAGGCTTTGCGGAACAGTGTAT GTCGGAATATGACCTGGATGGCTGGACCCGCCCGGACCTGATTAACCCGGGCACGTAA
CYP199A2 (*Rhodopseudomonas palustris*)	ATGACGACCGCTCCGAGCCTGATGCCGGTTACGACGCCGTCTCAACATGGTGCTGGTGTGCCGCATCTGGGTATCGACCCGT TCGCACTGGATTATTTTGCAGACCCGTACCCGGAACAGGAAACGCTGCGTGAAGCGGGTCCGGTGGTTTATCTGGATAAATG GAACGTTTACGGCGTCGCCCGCTATGCAGAAGTGTACGCGGTTCTGAATGATCCGCTGACCTTTTGCAGCTCTCGTGGCGTGG GTCTGTCAGACTTCAAAAAAGAAAAACCGTGGCGCCCGCCGTCGCTGATTCTGGAAGCTGATCCGCCGGCACATACGCGTAC CCGTGCTGTCCTGTCAAAAGTGCTGTCGCCGGCGACCATGAAACGTCTGCGCGATGGTTTTGCGGCCGCAGCTGATGCCAAA ATCGACGAACTGCTGGCACGTGGCGGTAACATTGATGCTATCGCGGACCTGGCCGAAGCATATCCGCTGTCAGTTTTTCCGG ATGCCATGGGTCTGAAACAGGAAGGCCGCGAAAATCTGCTGCCGTACGCTGGTCTGGTCCTGAACGCATTCGGTCCGCCGAA TGAACTGCGTCAGAGCGCCATTGAACGCTCTGCACCGCATCAGGCGTATGTTGCGGAACAGTGCCAACGTCCGAACCTGGCA CCGGGCGGTTTTGGTGCATGTATTCACGCATTCTCCGATACGGGCGAAATCACCCCGGAAGAAGCTCCGCTGCTGGTGCGTA GTCTGCTGTCCGCCGGTCTGGACACCACGGTGAACGGTATCGCAGCAGCAGTTTACTGCCTGGCCCGCTTTCCGGATGAATTC GCTCGTCTGCGTGCGGACCCGAGCCTGGCCCGTAATGCATTTGAAGAAGCAGTTCGCTTCGAATCTCCGGTCCAGACGTTTTT CCGTACCACGACCCGCGATGTCGAACTGGCTGGTGCGACCATTGGCGAAGGTGAAAAAGTGCTGATGTTTCTGGGCAGCGCA AATCGTGACCCGCGTCGCTGGGATGACCCGGATCGTTATGACATCACGCGCAAAACCAGTGGTCATGTTGGCTTCGGTTCCG GCGTTCACATGTGTGTCGGTCAACTGGTGGCGCGTCTGGAAGGTGAAGTCGTGCTGGCTGCACTGGCACGTAAAGTGGCAGC AATTGAAATCGCAGGCCCGCTGAAACGCCGTTTTAACAATACCCTGCGTGGTCTGGAAAGCCTGCCGATTCAACTGACCCCG GCCTGA

163 **2.3 Construction of plasmids**

164 The genes encoding TAL and C3H were expressed in *E. coli* cells using the pETDuet-1,
165 pCDFDuet-1, pRSFDuet-1 and pKVS45 vectors (Table 2). The gene encoding CYP199A2
166 mutant was cloned in pCDFDuet-1 and CYP199A2 redox partners, *pdr* and *pux* genes,
167 were cloned in an operon in pKVS45 using restriction enzymes. All the primers used are
168 summarized in Table 4 ~~3~~. TAL and CYP199A2 were also cloned in an operon using
169 Phusion DNA plymerase (NEB) and overlap extension polymerase chain reaction
170 (PCR)[21]. The ribosome binding site (RBS) chosen was the same used in pETDuet-1,
171 pCDFDuet-1 and pRSFDuet-1. Reverse primers of TAL gene were overlapped with
172 forward primers of the CYP199A2 gene to introduce the RBS and spacer. Briefly, the TAL
173 and CYP199A2 genes were amplified and the overlapping strands of these intermediate
174 products hybridized in a subsequent PCR and were extended to generate the full-length
175 product amplified by flanking primers that included restriction enzyme sites for inserting
176 the operon into the plasmid.

177 All construction plasmids described were verified by colony PCR or digestion and
178 sequenced by Macrogen (Amsterdam, The Netherlands) or Genewiz (Cambridge, MA,
179 USA).

Table 4 3. Set of primers for PCR amplification (forward and reverse primers – FW and REV).

Primer name	Primer sequence[a]	Restriction enzyme
TAL_pET_pCDF_pRSF_FW	*GGCGCGCC*AA**ATG**GCTCCGCGTCCG	*Asc*I
TAL_pET_pCDF_pRSF_REV	*GCGGCCGC***TTA**TGCCAGCATTTTCAGCAG	*Not*I
TAL_pKVS45_FW	*CCTAGG*AAGGAGATATACC**ATG**GGCAGCAGCCATCACCATCATCACCACAGCCAGGCTCCGCGTCCG	*Avr*II
TAL_pKVS45_REV	*GGATCC***TTA**TGCCAGCATTTTC	*Bam*HI
C3H_pET_FW	*AGATCT*C**ATG**ACGATTACCTCTCCGGC	*Bgl*II
C3H_pET_REV	*CTCGAG*CGTGCCCGGGTTAATCAG	*Xho*I
C3H_pKVS45_FW	*CCTAGG*AAGGAGATATACC**ATG**GGCAGCAGCCATCACCATCATCACCACAGCCAGACGATTACCTCTCCGGCA	*Avr*II
C3H_pKVS45_REV	*GGATCC***TTA**CGTGCCCGGGTTAATCAG	*Bam*HI
C3H_pCDF_pRSF_FW	*GGCGCGCC*AA**ATG**ACGATTACCTCTCCGG	*Asc*I
C3H_pCDF_pRSF_REV	*AAGCTT***TTA**CGTGCCCGGGTTAATC	*Hind*III
CYP(+7aa)_pCDF_FW	*CATATG*CATCACCATCATCACCAC**ATG**ACGACCGCTCCGAGCCT	*Nde*I
CYP_pCDF_FW	*CATATG*CATCACCATCATCACCAC**ATG**CCGGTTACGACG	*Nde*I
CYP_pCDF_REV	*CTCGAG***TCA**GGCCGGGGTC	*Xho*I
TAL_op_FW	*GGATCC*A**ATG**GCTCCGCGTC	*Bam*HI
TAL_op_REV	CGT**CAT**GGTATATCTCCTT**TTA**TGCCAGCATTTTCAGC	-
CYP(+7aa)_op_FW	A**TAA**AAGGAGATATACC**ATG**ACGACCGCTC	-
CYP_op_REV	*AAGCTT***TCA**GGCCGGGGTC	*Hind*III
TAL_op2_REV	GG**CAT**GGTATATCTCCTT**TTA**TGCCAGCATTTTCAGC	-
CYP_op2_FW	A**TAA**AAGGAGATATACC**ATG**CCGGTTACGACG	-
Pdr_pKVS45_FW	*CCTAGG*AATAATTTTGTTTAACTTTAAGAAGGAGATATA**ATG**AACGCAAACGAC	*Nde*I
Pdr_pKVS45_REV	*GAGCTC***TCA**GGCACTACTCAGTTCAGC	*Sac*I
Pux_pKVS45_FW	*GGATCC*AATAATTTTGTTTAACTTTAAGAAGGAGATATA**ATG**CCCAGTATCACGTTCATTCTT	*Bam*HI
Pux_pKVS45_REV	*GCATGC***TCA**GACCTGACGATCCGGAAT	*Sph*I
Pdr_Pux_op_pET_FW	*GAATTC*A**ATG**AACGCAAACGACAAC	*Eco*RI
Pdr_Pux_op_pET_REV	*GATATC***TCA**GACCTGACGATCCG	*Eco*RV

[a]Start and stop codons in **bold**, occasionally the start codon is placed upstream of the His_6-tag sequence and no stop codon is included because of the presence of a Strep-tag; restriction sites in *italic*; His_6-tag underlined; The Ribosome Binding Site (RBS) and spacer are double underlined; In order for the sequence to remain in frame one or two bases were occasionally added between the restriction site and the gene start codon.

184 **2.4 Growth Conditions – *p*-Coumaric and Caffeic Acid Production**

185 *E. coli* cells for gene cloning, plasmid propagation, and inoculum preparation were grown
186 in LB medium at 37 ºC and 200 rpm.

187 For *p*-coumaric acid and/or caffeic acid production, cultures were grown at 37 ºC in 50 mL
188 LB to an optical density at 600 nm (OD_{600}) of 0.4. IPTG and/or aTc were added at the same
189 time (unless otherwise specified) at a final concentration of 1 mM and 100 ng/mL,
190 respectively, and the culture was incubated for 5 h at 26 ºC. The cells were harvested by
191 centrifugation, resuspended in 50 mL of modified M9 minimal salt medium containing (per
192 liter): glucose (40 g), Na_2HPO_4 (6 g), KH_2PO_4 (3 g), NH_4Cl (1g), NaCl (0.5 g), $CaCl_2$ (17
193 mg), $MgSO_4$ (58 mg), thiamine (340 mg) and $CaC0_3$ (5 g) (to control the pH), and
194 incubated at 26 ºC for 63 h. Trace elements [$FeCl_3$ (54 mg), $ZnCl_2$ (4 mg), $CoCl_2$ (4 mg),
195 $NaMoO_4$ (4 mg), $CuCl_2$ (2 mg) and H_2BO_3 (1 mg)] and vitamins [riboflavin (0.84 mg),
196 folic acid (0.084 mg), nicotinic acid (12.2 mg), pyridoxine (2.8 mg), biotin (0.12 mg) and
197 pantothenic acid (10.8 mg)] were also supplemented to the M9 medium. Depending on the
198 plasmid(s) present in the strain, 100 μg/mL of ampicillin, 100 μg/mL of spectinomycin
199 and/or 50 μg/mL of kanamycin were added. IPTG and/or aTc and substrates were added at
200 time 0 of induction in M9 medium (unless otherwise stated): tyrosine, 3 mM or *p*-coumaric
201 acid, 2 mM. Samples of the supernatant (1.5 mL) were collected at time 0 and after 15, 24,
202 43 and 63 h. All the experiments were done in triplicate and analyzed by high-performance
203 liquid chromatography (HPLC).

204

205

206 **2.5 HPLC analysis of the products**

207 HPLC analysis was used to quantify *p*-coumaric acid and caffeic acid using a HPLC system
208 from Jasco (Easton, MD, USA) (PU-2080 Plus Pump unit, LG-2080-02 Ternary Gradient
209 unit, a DG-2080-53 3-Line Degasser unit, a UV-2075 Plus Intelligent UV/VIS Detector
210 unit and AS-2057 Plus Intelligent Sampler unit) and a Grace Alltech Platinum EPS C18
211 column (3µm, 150 mm × 4.6 mm) (Grace, Columbia, MD, USA). Mobile phases A and B
212 were composed of water (0.1% trifluoroacetic acid) and acetonitrile, respectively. The
213 following gradient was used at a flow rate of 1 mL/min: 10 - 20% acetonitrile (mobile
214 phase B) for 17 min. Quantification was based on the peak areas of absorbance at 275 nm
215 (tyrosine) and 310 nm (*p*-coumaric acid and caffeic acid). The retention times of tyrosine,
216 *p*-coumaric acid and caffeic acid were 3.3, 8.0 and 11.8 min, respectively.

217 **2.6 Protein Analysis**

218 *E. coli* K-12 MG1655(DE3) cells harboring pETDuet-1, pETDuet_TAL, pETDuet_C3H
219 and pETDuet_TAL_C3H were grown in LB at 37 ºC to an OD_{600} of 0.6. IPTG was added at
220 a final concentration of 1 mM, and the culture was incubated for 24 h. Samples (10 mL
221 culture medium) were taken at times 0, 5 and 24 h. Samples were centrifuged and the cells
222 were resuspended in phosphate-buffered saline (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM
223 Na_2PO_4, 1.8 mM KH_2PO_4, pH 7.4) and further disrupted by sonication on ice for 3 min.
224 After centrifugation the protein concentration from the resulting supernatant was
225 determined using Protein Assay Dye Reagent Concentrate (BioRad, Hercules, CA, USA)
226 with bovine serum albumin (BSA) (NEB) as a standard. The expression levels of TAL and
227 C3H were evaluated through sodium dodecyl sulfate polyacrylamide gel electrophoretic
228 (SDS–PAGE). Fifteen to 20 µg of total protein were loaded onto a 4-20 % Mini-

PROTEAN® TGX™ Precast Gels (BioRad). The protein marker used was Precision Plus Protein™ Unstained (BioRad). For gel staining, Bio-Safe Coomassie Stain (BioRad) was used.

3. Results and Discussion

3.1 Selection of the appropriate enzyme sources

TAL was chosen from the red yeast *R. glutinis*, since it was reported to have the highest TAL activity and a low PAL/TAL catalytic activity ratio [22-24]. TAL prefers tyrosine and is therefore favored over PAL and C4H, both of which are required to begin the pathway from phenylalanine (Fig. 1). This decreases the pathway number of steps and at the same time eliminates the issues associated with the use of C4H. C4H is a P450-dependent monooxygenase and the functional expression of plant P450 enzymes is always hard to achieve in bacteria, which is mainly due to the absence of cytochrome P450 reductases (CPRs)/redox partners in *E. coli* needed for electron transfer, and to the absence of endoplasmatic reticulum which prevents the efficient translational of the membrane signal modules of microsomal P450 enzymes [25]. C3H from *S. espanaensis* was chosen as it is one of the rare cytochrome P450 enzymes that was successfully expressed in *E. coli* [15] (Table 1). CYP199A2 from *R. palustris* was chosen as it was effectively used in previous studies to achieve one of the highest caffeic acid production from *p*-coumaric acid [8].

252 **3.2 Production of *p*-coumaric acid from tyrosine using TAL**

253 In order to study *p*-coumaric acid production from tyrosine, codon-optimized TAL was
254 cloned in pETDuet-1, pCDFDuet-1, pRSFDuet-1 and pKVS45. Fig. 2 illustrates *p*-
255 coumaric acid production from 3 mM of tyrosine. *p*-Coumaric acid was also produced from
256 endogenous tyrosine in the presence of pCDFDuet_TAL, but in a very low quantity
257 compared to conditions with tyrosine supplemented. Since the endogenous tyrosine is not
258 enough to produce *p*-coumaric acid in high concentrations, two strategies can be used,
259 namely supplement the medium with tyrosine or engineer *E. coli* to overproduce tyrosine
260 from glucose. *p*-Coumaric acid production using TAL was found to be highly dependent on
261 the plasmid chosen. The highest production obtained was 2.62 mM (472 mg/L) and it was
262 achieved with pRSFDuet-1, which is a high copy number plasmid. In this case, the product
263 yield was 0.87 (mol *p*-coumaric acid/ mol tyrosine) and the remaining tyrosine was
264 presumably used for growth and primary metabolism since no tyrosine was detected by
265 HPLC. Santos et al. [26] used a codon-optimized *R. glutinis* TAL with a 80% similarity to
266 the one used in this study. Although both plasmids (pETDuet-1 and pTrcHis2B) have a
267 pBR322 origin, they have different promoters, with the T7 system being stronger than the
268 *trc* promoter. The *p*-coumaric acid production obtained using pETDuet_TAL (the lowest
269 production in the current study) was higher than that reported by Santos et al. [26] after 72
270 h using pTrcHis2B (104 mg/L – 0.61 mM). They obtained a product yield of 0.23 (mol *p*-
271 coumaric acid / mol tyrosine) in a *E. coli* K12 MG1655(DE3) strain. From analyzing these
272 results it can be concluded that the *R. glutinis* TAL used in our study and the expression
273 systems chosen are a very good option to produce *p*-coumaric acid.

274

Fig. 2B shows the production of *p*-coumaric acid in the strain harboring pRSFDuet_TAL over time (63 h). The results showed that the *p*-coumaric concentration increases along time, and this trend was also observed when using other plasmid constructions (*data not shown*). Although 63 hours of incubation is a long time, it is not unusual to prolong the fermentations more than 24 h in the caffeic acid production to obtain higher titers and yields, as can be observed in Table 1. Albeit the incubations can be performed at higher temperatures, the production would be lower since our preliminary results showed that using TAL and the other enzymes used in this study at 30 ºC and 37 ºC leads to lower titers than 26 ºC (*data not shown*).

3.3 Production of caffeic acid from *p*-coumaric acid using C3H

Caffeic acid was produced from *p*-coumaric acid using *S. espanaensis* C3H (Fig. 3). C3H, although being a plant cytochrome P450 enzyme, has been successfully expressed in *E. coli* [15, 17, 18]. As previously shown in the *p*-coumaric acid production from tyrosine (Fig. 2), the highest caffeic acid titer was also obtained with the high copy number plasmid pRSFDuet-1. The titer obtained was around 0.93 mM (168 mg/L), corresponding to a product yield of 0.47 (mol caffeic acid / mol *p*-coumaric acid).

3.4 Production of caffeic acid from tyrosine using TAL and C3H

In a second phase of the current study, TAL and C3H were combined to produce caffeic acid from 3 mM of tyrosine (Fig. 4). In a first approach, the genes were cloned together in pETDuet-1. When these genes were combined, the production of caffeic acid increased 170 %. This increase is probably due to the fact that in this case the *p*-coumaric concentration in

the medium is lower than when it is added directly, thus reducing the toxic effect to the cells as previously described in the literature [9, 10, 27-30]. To confirm that the addition of *p*-coumaric acid had a toxic effect to the cells, *p*-coumaric acid and tyrosine were added separately to *E. coli* harbouring pETDuet-1 and pETDuet_TAL_C3H. When *p*-coumaric acid was added, the *E. coli* cultures had a lower growth rate and the OD_{600} was 10-11% lower than when tyrosine was added (*data not shown*). Moreover, the protein expression was found to be lower. When only TAL is cloned in the pETDuet-1 plasmid the protein band in the SDS gel could be clearly visualized (76.34 kDa) (*data not shown*) ~~(Fig. S1)~~. However, when TAL was combined with C3H, protein production was not observed maybe due to the metabolic burden imposed on the host cells by the expression of the two proteins in the same plasmid. Although in small amounts, a band around 56.33 kDa corresponding to C3H production could be observed.

Cloning TAL in pCDFDuet-1 and C3H in pETDuet-1 led to results very similar to the experiments in which these enzymes were used alone to produce *p*-coumaric acid or caffeic acid, respectively (Fig. 2A and Fig. 3). This occurs because the expression of TAL when alone in the plasmid is very high ~~(Fig. S1)~~, leading to high production of *p*-coumaric acid. Consequently, caffeic acid amounts are comparable to the ones obtained when *p*-coumaric acid is added as substrate. Similar results were observed with the combination pETDuet_TAL and pCDFDuet_C3H. The production of *p*-coumaric acid was 1.28 mM and caffeic acid was 0.77 mM. Since caffeic acid is produced from *p*-coumaric, this suggests that in total more than 2 mM of *p*-coumaric acid was produced. When only pETDuet_TAL was used (Fig. 2A), the maximum *p*-coumaric acid production obtained was 0.9 mM, which demonstrates that tyrosine can be converted faster if *p*-coumaric acid is being converted to

the next product. Again we observed that caffeic acid production does not increase (compared to the cases when only C3H is used and the substrate is *p*-coumaric acid) since the *p*-coumaric acid concentration in the medium is still very high. The combinations pKVS45_TAL/ pCDFDuet_C3H, pCDFDuet_TAL/ pKVS45_C3H and pRSFDuet_TAL/ pCDFDuet_C3H showed very similar results.

The combination pCDFDuet_TAL/pRSFDuet_C3H is very interesting as almost all the *p*-coumaric acid produced is converted to caffeic acid. In the end, only around 3 μM of *p*-coumaric acid is detected by HPLC and the concentration in time was never higher than 0.15 mM, meaning that almost all the *p*-coumaric acid being produced was being converted by C3H to caffeic acid. A final caffeic acid concentration of around 1 mM (180 mg/L) was obtained. This combination enabled the highest yield from tyrosine with an additional benefit of having in the end a very low amount of the intermediate *p*-coumaric acid, contrary to what was observed in the other cases.

Zhang and Stephanopoulos [18] used codon optimized TAL from *R. glutinis* and C3H from *S. espanaensis* to produce caffeic acid. Also, the authors used a tyrosine over producing strain and no tyrosine limitation was observed during the caffeic acid production. After several medium optimizations, the highest titer they reported after 72 h was 106 mg/L (Table 1), which is 1.7 times lower than the maximum titer obtained in the current study after 63 h. Although those authors used the same plasmid to carry C3H (pRSFDuet-1), they used the pTrcHis2B plasmid to carry TAL, which was previously described by Santos et al. [26] and that we concluded it is not the best plasmid for this gene and to produce *p*-coumaric acid (See Production of p-coumaric acid from tyrosine using TAL section). Our

study reports the highest titer of caffeic acid produced so far using the combination of TAL and C3H genes.

3.5 Production of caffeic acid from *p*-coumaric acid using CYP199A2

Until the recent studies published using 4HPA3H [9, 10], CYP199A2 gene from *R. palustris* was reported to produce the highest amounts of caffeic acid from *p*-coumaric acid [8]. The results obtained in our study with CYP199A2 are summarized in Fig. 5. CYP199A2 was cloned in pCDFDuet-1 and its redox partners, Pdr and Pux, were expressed in another plasmid (pKVS45 or pETDuet-1) as part of an operon as described before [31, 32]. The first results obtained with CYP199A2 in the caffeic acid production were surprisingly low as compared to the Furuya et al. [8] report. After analyzing the CYP199A2 DNA sequence it was verified that those authors did not use the first 21 bp of CYP199A2 to clone the gene. This decision was based on software results that annotated the 8th CYP199A2 amino acid (GTG) as a start codon before the sequence was published (T. Furuya, personal communication). We also confirmed that, for example, EasyGene 1.2b Server [33, 34] identified the 8th amino acid as the start codon. Based on this, we cloned CYP199A2 without the first 7 amino acids. The production using this new CYP199A2 increased considerably the caffeic acid production compared to the original clone (CYP199A2(+7aa)). Additionally, it was found that the production was more than 2.7 times higher when the inducer of pKVS45_Pdr_Pux_op, aTc, was added 2.5 h after addition of IPTG (to induce CYP199A2 ~~TAL~~ expression). The delay of induction can alleviate ~~d~~ the metabolic burden of several plasmids [26, 35]. Since pKVS45 has CYP199A2 redox partners (Pdr and Pux) and these proteins are only needed after CYP199A2 is present in a

366 significant concentration to support its catalytic activity (NADPH- and O_2- dependent
367 hydroxylation reactions), their expression can be delayed. The results obtained with
368 CYP199A2 in this case were 1.8 times better than the one obtained with C3H in pCDFDuet
369 adding *p*-coumaric acid (Fig. 3). Nevertheless, to improve the yield and taking into
370 consideration that the addition of a high concentration of *p*-coumaric acid can have a
371 detrimental effect on the caffeic acid production, as discussed before, a different feeding
372 system was tested: 1 mM of *p*-coumaric acid was added at time 0 of induction in M9
373 medium and 0.5 mM added at 5 h and 24 h. Caffeic acid production increased to 1.72 mM
374 (310 mg/L) using this three step feeding. This approach was successfully demonstrated
375 before for the production of caffeic acid [9, 10] and allowed us to obtain a product yield of
376 0.86 (mol caffeic acid / mol *p*-coumaric acid) after 63 h. Furuya et al. [8] obtained a
377 maximum yield of 0.75 (mol caffeic acid / mol *p*-coumaric acid) after 24 h using glycerol
378 as ~~after testing different~~ energy source ~~s~~. When the authors used glucose, the maximum
379 yield obtained after 24 h was around 0.46 (mol caffeic acid / mol *p*-coumaric acid), thus
380 very similar to the one obtained in the current study after 24 h - 0.47 (mol caffeic acid / mol
381 *p*-coumaric acid), which suggests that glycerol is a more effective energy source to
382 regenerate NADH from NAD^+ [8].

383 Since pKVS45 time of induction with aTc seems to limit caffeic acid production, pETDuet-
384 1, that is induced using IPTG similarly to pCDFDuet-1, was chosen to clone the
385 CYP199A2 redox partners and evaluate if the caffeic acid production could be increased.
386 However, the results obtained with pETDuet-1 were very similar to the ones obtained with
387 pKVS45 induced at time 0. This result is not surprising since pKVS45 alone, or combined

388 with pCDFDuet-1, gave better overall results using TAL and C3H than the pETDuet-1
389 plasmid (Fig. 2-4).

390

391 **3.6 Production of caffeic acid from tyrosine using TAL and CYP199A2**

392 To produce caffeic acid from tyrosine, TAL and CYP199A2 were combined using different
393 approaches, namely together in the same plasmid but in different MCSs (TAL in MCS1 and
394 CYP199A2 in MCS2), or in an operon in MCS1 (Fig. 6). Both CYP199A2 sequences were
395 tested and again it was concluded that CYP199A2 starting at the 8th amino acid provides
396 better results. The low caffeic acid concentration obtained was due to the addition of aTc at
397 time zero of induction. However, even with a low concentration, it is possible to conclude
398 that the genes work better when cloned in different MCSs than when cloned in an operon.
399 Thus, this approach was chosen to proceed with other tests including pKVS45 later
400 induction with aTc; three phases of *p*-coumaric acid addition; and the use of another
401 plasmid to carry CYP199A2 redox partners (pETDuet-1). The results obtained were very
402 similar to the ones found with only CYP199A2 in the plasmid (Fig. 5). The expression of
403 TAL in a different plasmid (pRSFDuet-1) improved caffeic acid production. The highest
404 caffeic acid production was 1.56 mM (280 mg/L). According to these results, CYP199A2
405 with its redox partners seems to be a better option than C3H to produce caffeic acid from
406 tyrosine or *p*-coumaric acid.

407 Based on the above discussion, we believe that it would be advisable to, in the future, clone
408 CYP199A2 with its redox partners in different plasmids and combine them with TAL to
409 confirm if caffeic acid production can be improved without a high accumulation of the
410 intermediate as its accumulation may result in suboptimal production titers. Although no

kinetic parameters were determined in our study, it was possible to observe that over time the conversion of tyrosine to *p*-coumaric acid was faster than the conversion of *p*-coumaric acid to caffeic acid. The faster conversion of tyrosine to *p*-coumaric acid leads to the accumulation of *p*-coumaric acid which represents a drawback to the caffeic acid production since its toxicity leads to an even more pronounced decrease of the production. The kinetic parameters of the enzymes used in this study should be determined, especially regarding the C3H, since TAL and CYP199A2 kinetic parameters were characterized by Xue et al. [22] and Furuya et al. [8], respectively. Furthermore, it will be important to evaluate a combination of plamids that allows a fine-tuned production like the one observed when TAL was cloned in pCDFDuet-1 and C3H in pRSFDuet-1, where no *p*-coumaric acid was accumulated (Fig.4). Also, to improve the yield and to avoid the need of two separate stages of cultivation for biomass/protein generation and caffeic acid production, the use of potassium phosphate buffer [8, 9] or MOPS [18, 26] with glucose or glycerol should be tested. The use of M9 minimal medium without the production of biomass and protein in LB does not allow obtaining productions as high as the ones obtained by first using LB and then transferring the cells to M9 (*data not shown*).

Caffeic acid has for a long time been recognized for its therapeutic properties, which makes it an attractive target for metabolic engineering and synthetic biology. We have successfully designed a pathway for the production of caffeic acid via metabolic engineering approaches in *E. coli*. We tested different genetic arrangements with two (or four) genes to balance the expression of the enzymes and achieve an optimized performance, and we obtained significantly different levels of caffeic acid productions. The layout of genes and operons in the plasmid, as well as the use of different plasmids had an

enormous impact on gene expression. In addition, codon-optimization when expressing heterologous genes in *E. coli* was considered to improve gene expression. The repeated addition of the substrate and the delay in the induction of protein expression also led to an increase of the titers by decreasing the toxicity of *p*-coumaric acid and the metabolic burden of heterologous protein expression.

In conclusion, the caffeic acid is a phenylpropanoic acid and this pathway can be further used to produce other products of the phenylpropanoid pathway from tyrosine like flavonoids, stilbenoids, isoflavonoids and curcuminoids. Until now the biosynthesis of these compounds is in the range of the titers obtained in this study for *p*-coumaric acid and caffeic acid or much lower. Therefore, the strategy of adding *p*-coumaric acid at concentrations as high as 20 mM, concentrations used in some studies for the caffeic acid production (Table 1), was considered unreasonable, especially knowing that high concentrations of *p*-coumaric acid are toxic to the cells. For the production system to be even more economically viable, the use of a tyrosine overproducing strain should be considered.

Competing Interests

The authors declare no competing interests.

Abbreviations

456 TAL: Tyrosine Ammonia Lyase; C3H: 4-coumarate 3-hydroxylase; PAL: Phenylalanine
457 Ammonia Lyase; C4H: cinnamate-4-hydroxylase; 4HPA3H: hydroxyphenylacetate 3-
458 hydroxylase 4CL: 4-coumarate-CoA ligase; MCS: Multiple Cloning Sites; PCR:
459 Polymerase Chain Reaction; RBS: Ribosome Binding Site; HPLC: High-Performance
460 Liquid Chromatography; SDS-PAGE: Sodium Dodecyl Sulfate - Polyacrylamide Gel
461 Electrophoretic (SDS–PAGE).

462

463

464 ## Acknowledgments

465 This work was supported by FEDER funds through the COMPETE and ON2 program and
466 through National funds of FCT in the scope of the project FCOMP-01-0124-FEDER-
467 027462, PEst-OE/EQB/LA0023/2013, NORTE-07-0124-FEDER-000028 and NORTE-07-
468 0124-FEDER-000027. Financial support for this work was provided by FCT grant SFRH /
469 BD / 51187 / 2010 and SYNBIOBACTHER project (PTDC/EBB-BIO/102863/2008).

470

471 ## References

472 [1] Mori H, Iwahashi H. Antioxidant activity of caffeic acid through a novel mechanism under uva
473 irradiation. Journal of Clinical Biochemistry and Nutrition. 2009;45:49-55.
474 [2] Chao P, Hsu C, Yin M. Anti-inflammatory and anti-coagulatory activities of caffeic acid and
475 ellagic acid in cardiac tissue of diabetic mice. Nutrition and Metabolism. 2009;6:1-8.
476 [3] Prasad NR, Karthikeyan A, Karthikeyan S, Reddy BV. Inhibitory effect of caffeic acid on cancer
477 cell proliferation by oxidative mechanism in human HT-1080 fibrosarcoma cell line. Mol Cell
478 Biochem. 2011;349:11-9.
479 [4] Ikeda K, Tsujimoto K, Uozaki M, Nishide M, Suzuki Y, Koyama AH, et al. Inhibition of
480 multiplication of herpes simplex virus by caffeic acid. International Journal of Molecular Medicine.
481 2011;28:595-8.
482 [5] Jung UJ, Lee M-K, Park YB, Jeon S-M, Choi M-S. Antihyperglycemic and antioxidant properties of
483 caffeic acid in db/db mice. J Pharmacol Exp Ther. 2006;318:476-83.
484 [6] Takeda H, Tsuji M, Inazu M, Egashira T, Matsumiya T. Rosmarinic acid and caffeic acid produce
485 antidepressive-like effect in the forced swimming test in mice. Eur J Pharmacol. 2002;449:261-7.

[7] Xing Y, Peng H-y, Zhang M-x, Li X, Zeng W-w, Yang X-e. Caffeic acid product from the highly copper-tolerant plant *Elsholtzia splendens* post-phytoremediation: its extraction, purification, and identification. Journal of Zhejiang University Science B. 2012;13:487-93.
[8] Furuya T, Arai Y, Kino K. Biotechnological production of caffeic acid by bacterial cytochrome P450 CYP199A2. Appl Environ Microbiol. 2012;78:6087-94.
[9] Furuya T, Kino K. Catalytic activity of the two-component flavin-dependent monooxygenase from *Pseudomonas aeruginosa* toward cinnamic acid derivatives. Appl Microbiol Biotechnol. 2013;98:1145-54.
[10] Huang Q, Lin Y, Yan Y. Caffeic acid production enhancement by engineering a phenylalanine over-producing *Escherichia coli* strain. Biotechnol Bioeng. 2013;110:3188-96.
[11] Miyahisa I, Kaneko M, Funa N, Kawasaki H, Kojima H, Ohnishi Y, et al. Efficient production of (2S)-flavanones by *Escherichia coli* containing an artificial biosynthetic gene cluster. Appl Microbiol Biotechnol. 2005;68:498-504.
[12] Hwang EI, Kaneko M, Ohnishi Y, Horinouchi S. Production of plant-specific flavanones by *Escherichia coli* containing an artificial gene cluster. Appl Environ Microbiol. 2003;69:2699-706.
[13] Watts KT, Lee PC, Schmidt-Dannert C. Exploring recombinant flavonoid biosynthesis in metabolically engineered *Escherichia coli*. ChemBioChem. 2004;5:500-7.
[14] Berner M, Krug D, Bihlmaier C, Vente A, Müller R, Bechthold A. Genes and enzymes involved in caffeic acid biosynthesis in the actinomycete *Saccharothrix espanaensis*. J Bacteriol. 2006;188:2666-73.
[15] Choi O, Wu C-Z, Kang SY, Ahn JS, Uhm T-B, Hong Y-S. Biosynthesis of plant-specific phenylpropanoids by construction of an artificial biosynthetic pathway in *Escherichia coli*. Journal of Industrial Microbiology & Biotechnology. 2011;38:1657-65.
[16] Lin Y, Yan Y. Biosynthesis of caffeic acid in *Escherichia coli* using its endogenous hydroxylase complex. Microbial Cell Factories. 2012;11:1-9.
[17] Kang S-Y, Choi O, Lee JK, Hwang BY, Uhm T-B, Hong Y-S. Artificial biosynthesis of phenylpropanoic acids in a tyrosine overproducing *Escherichia coli* strain. Microbial Cell Factories. 2012;11:1-9.
[18] Zhang H, Stephanopoulos G. Engineering *E. coli* for caffeic acid biosynthesis from renewable sugars. Appl Microbiol Biotechnol. 2013;97:3333–41.
[19] Nielsen DR, Yoon SH, Yuan CJ, Prather KL. Metabolic engineering of acetoin and meso-2, 3-butanediol biosynthesis in *E. coli*. Biotechnol J. 2010;5:274-84.
[20] Solomon KV, Sanders TM, Prather KL. A dynamic metabolite valve for the control of central carbon metabolism. Metab Eng. 2012;14:661-71.
[21] Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. Gene. 1989;77:61-8.
[22] Xue Z, McCluskey M, Cantera K, Ben-Bassat A, Sariaslani FS, Huang L. Improved production of *p*-hydroxycinnamic acid from tyrosine using a novel thermostable phenylalanine/tyrosine ammonia lyase enzyme. Enzyme Microb Technol. 2007;42:58-64.
[23] Gatenby AA, Sariaslani S, Tang X-S, Qi WW, Vannelli T. Bioproduction of para-hydroxycinnamic acid. US Patent 6,368,837 ed: US Patent 6,368,837; 2002.
[24] Vannelli T, Wei Qi W, Sweigard J, Gatenby AA, Sariaslani FS. Production of *p*-hydroxycinnamic acid from glucose in *Saccharomyces cerevisiae* and *Escherichia coli* by expression of heterologous genes from plants and fungi. Metab Eng. 2007;9:142-51.
[25] Chapple C. Molecular-genetic analysis of plant cytochrome P450-dependent monooxygenases. Annual Review of Plant Biology. 1998;49:311-43.
[26] Santos CNS, Koffas M, Stephanopoulos G. Optimization of a heterologous pathway for the production of flavonoids from glucose. Metab Eng. 2011;13:392-400.

[27] Barthelmebs L, Diviès C, Cavin J-F. Expression in *Escherichia coli* of native and chimeric phenolic acid decarboxylases with modified enzymatic activities and method for screening recombinant *E. coli* strains expressing these enzymes. Appl Environ Microbiol. 2001;67:1063-9.
[28] Jung D-H, Choi W, Choi K-Y, Jung E, Yun H, Kazlauskas RJ, et al. Bioconversion of p-coumaric acid to p-hydroxystyrene using phenolic acid decarboxylase from *B. amyloliquefaciens* in biphasic reaction system. Appl Microbiol Biotechnol. 2013;97:1501-11.
[29] Shin S-Y, Han NS, Park Y-C, Kim M-D, Seo J-H. Production of resveratrol from *p*-coumaric acid in recombinant *Saccharomyces cerevisiae* expressing 4-coumarate: coenzyme A ligase and stilbene synthase genes. Enzyme Microb Technol. 2011;48:48-53.
[30] Watts KT, Lee PC, Schmidt-Dannert C. Biosynthesis of plant-specific stilbene polyketides in metabolically engineered *Escherichia coli*. BMC Biotechnol. 2006;6:1-12.
[31] Furuya T, Kino K. Discovery of 2-naphthoic acid monooxygenases by genome mining and their use as biocatalysts. ChemSusChem. 2009;2:645-9.
[32] Furuya T, Kino K. Regioselective oxidation of indole-and quinolinecarboxylic acids by cytochrome P450 CYP199A2. Appl Microbiol Biotechnol. 2010;85:1861-8.
[33] Larsen TS, Krogh A. EasyGene–a prokaryotic gene finder that ranks ORFs by statistical significance. BMC Bioinformatics. 2003;4:1-15.
[34] Nielsen P, Krogh A. Large-scale prokaryotic gene prediction and comparison to genome annotation. Bioinformatics. 2005;21:4322-9.
[35] Wu J, Du G, Zhou J, Chen J. Metabolic engineering of *Escherichia coli* for (2S)-pinocembrin production from glucose by a modular metabolic strategy. Metab Eng. 2013;16:48-55.

Figure Captions:

Fig. 1. Artificial caffeic acid biosynthetic pathway. The dashed box represents the caffeic acid pathway in plants. Inside and outside the box some strategies used to produce caffeic acid in *E. coli* are illustrated. TAL: Tyrosine ammonia lyase; C3H: 4-coumarate 3-hydroxylase; 4HPA3H: hydroxyphenylacetate 3-hydroxylase; 4CL: 4-coumarate-CoA ligase.

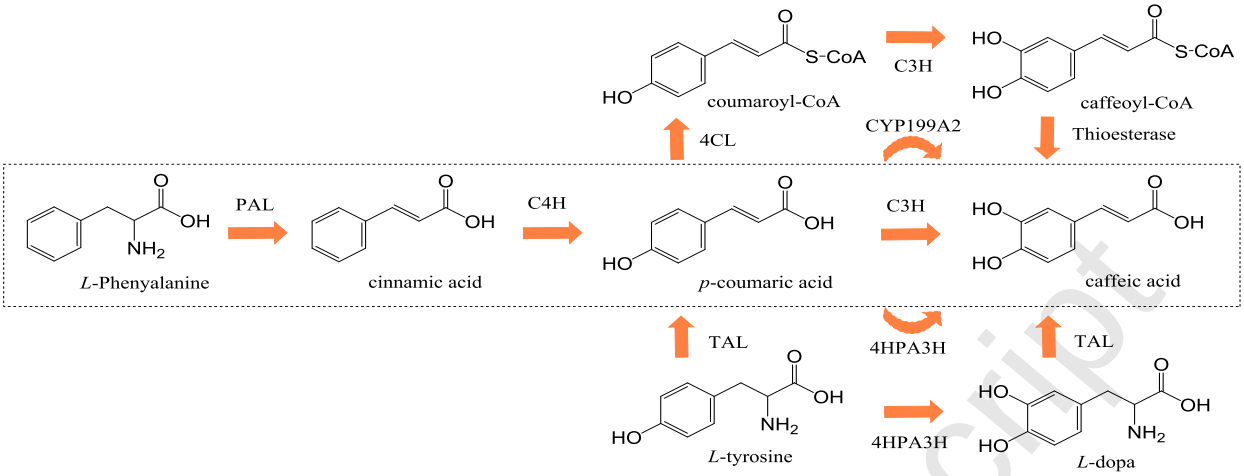
Fig. 2. Production of *p*-coumaric acid using TAL from *Rhodotorula glutinis* using different plasmids after 63 h (A) and using pRSFDuet_TAL during 63 h in M9 medium (B). (*) no substrate added. TAL: Tyrosine ammonia lyase.

Fig. 3. Production of caffeic acid using opC3H from *Saccharothrix espanaensis.* C3H: 4-coumarate 3-hydroxylase.

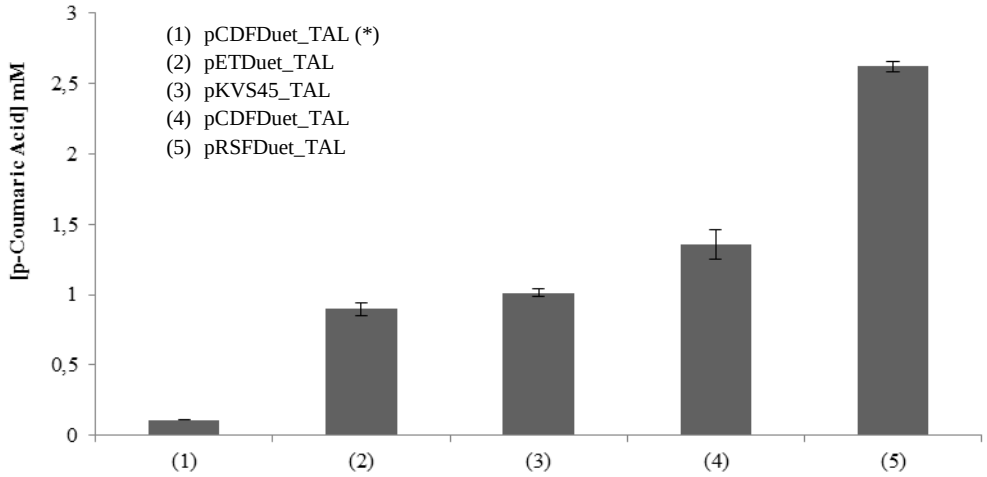
568 **Fig. 4.** Production of caffeic acid using opTAL from *Rhodotorula glutinis* and opC3H from
569 *Saccharothrix espanaensis*. TAL: Tyrosine ammonia lyase; C3H: 4-coumarate 3-
570 hydroxylase.

571 **Fig. 5.** Production of caffeic acid from *p*-coumaric acid using CYP199A2 from
572 *Rhodotorula palustris.* CYP: cytochrome P450 CYP199A2.

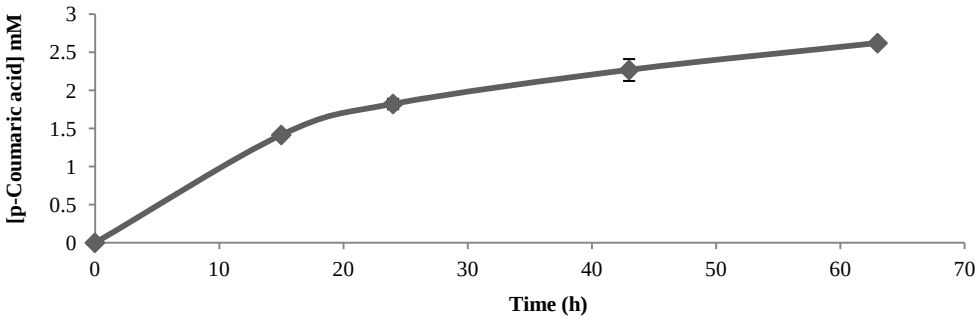
573 **Fig. 6.** Production of caffeic acid using TAL from *Rhodotorula glutinis* and CYP199A2
574 from *Rhodopseudomonas palustris*. TAL: Tyrosine ammonia lyase; CYP: cytochrome P450
575 CYP199A2.

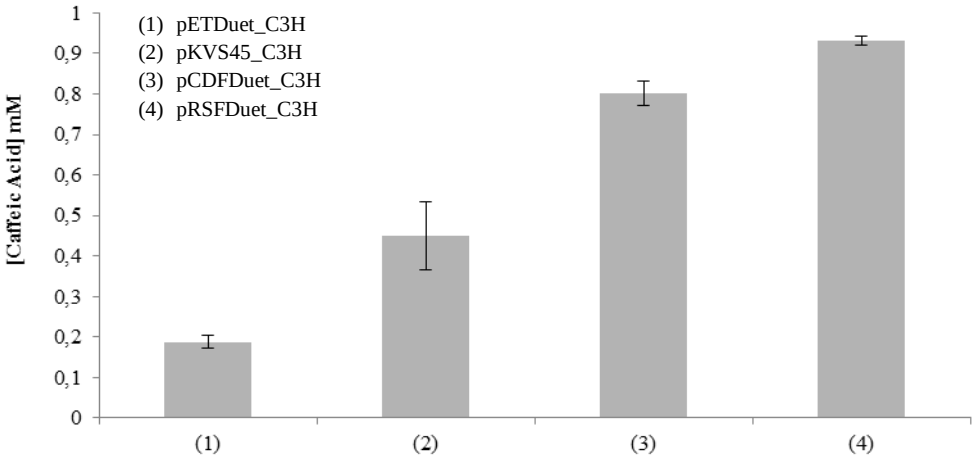


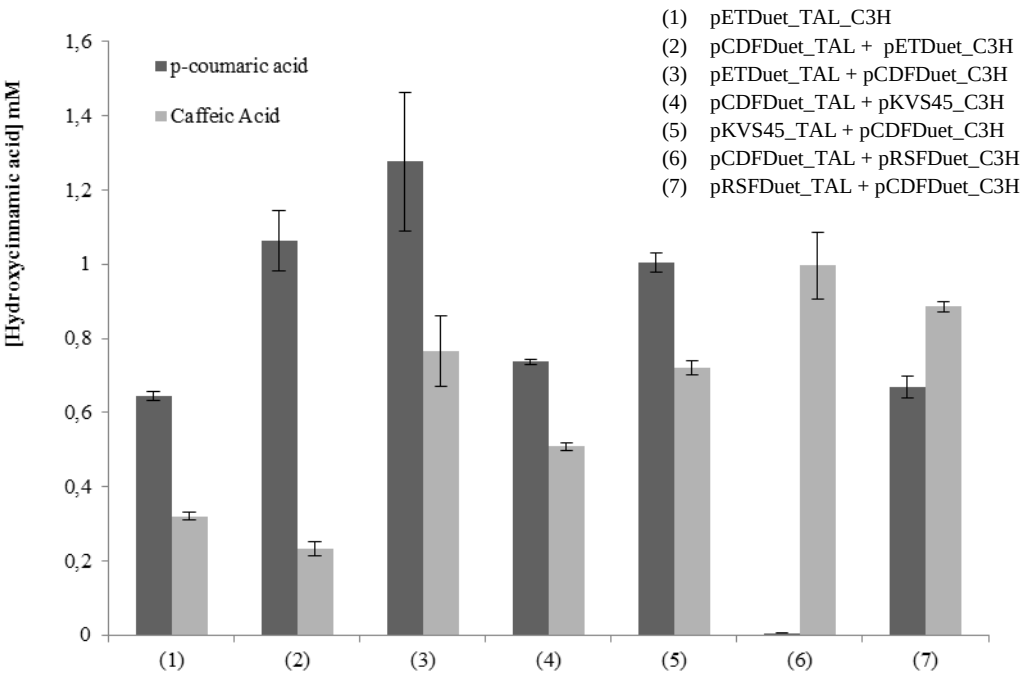
A

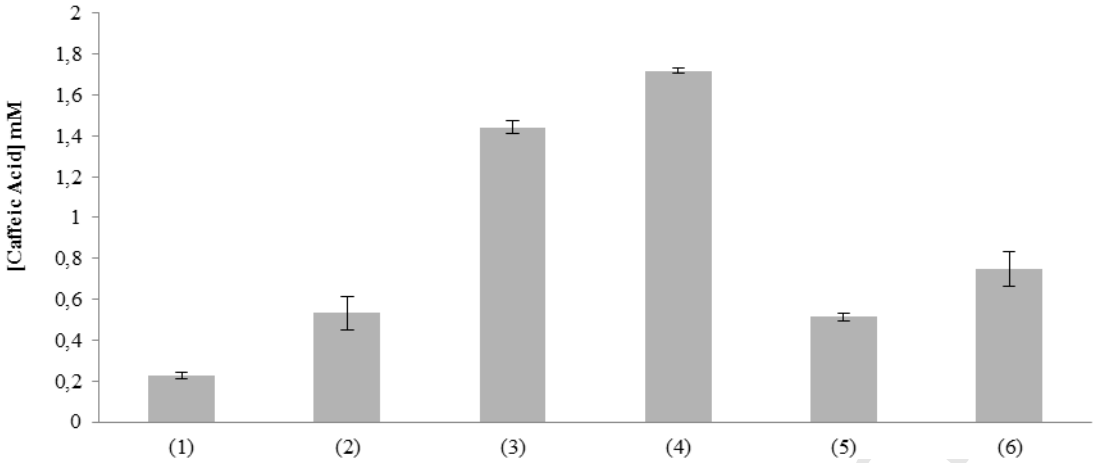


B

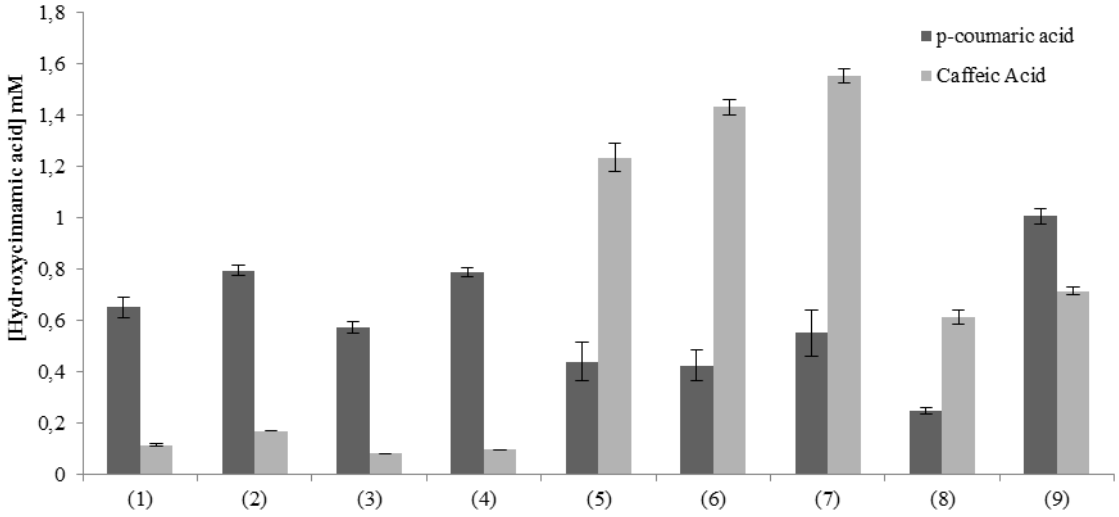








(1) pCDFDuet_CYP(+7aa) + pKVS45_PdrPux_op
(2) pCDFDuet_CYP + pKVS45_PdrPux_op
(3) pCDFDuet_CYP + pKVS45_PdrPux_op (aTc added 2,5 h later)
(4) pCDFDuet_CYP + pKVS45_PdrPux_op (aTc added 2,5 h later; *p*-coumaric acid added several times)
(5) pCDFDuet_CYP + pETDuet_PdrPux_op
(6) pCDFDuet_CYP + pETDuet_PdrPux_op (*p*-coumaric acid added several times)



(1) pCDFDuet_TAL_CYP(+7aa) + pKVS45_Pdr_Pux_op
(2) pCDFDuet_TAL_CYP + pKVS45_Pdr_Pux_op
(3) pCDFDuetTAL_CYP(+7aa)_op + pKVS45_PdrPux_op
(4) pCDFDuet_TAL_CYP_op + pKVS45_Pdr_Pux_op
(5) pCDFDuet_TAL_CYP + pKVS45_Pdr_Pux_op (aTc added 2,5 h later)
(6) pCDFDuet_TAL_CYP + pKVS45_Pdr_Pux_op (aTc added 2,5 h later and *p*-coumaric acid added several times)
(7) pRSFDuet_TAL + pCDFDuet_CYP + pKVS45_Pdr_Pux_op (aTc added 2,5 h later)
(8) pCDFDuet_TAL_CYP + pETduet_Pdr_Pux_op
(9) pRSFDuet_TAL + pCDFDuet_CYP + pETDuet_Pdr_Pux_op